

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:06:31 ; Search time 50 Seconds  
(without alignments)  
1601.599 Million cell updates/sec

Title: US-10-027-000-2

Perfect score: 4391  
Sequence: 1 MADIDVEALIKRLTLAEKVD.....DGVLRGKFRVGETYMGSGV. 833

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR\_73:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2052	46.7	845	1	GLVK
2	1445.5	32.9	818	2	A42292
3	1145	26.1	754	2	S04381
4	959	21.8	721	2	C72428
5	957	21.8	859	2	T35785
6	946.5	21.6	762	2	H87466
7	939.5	21.4	793	2	AF0341
8	874	19.9	758	2	D87369
9	860	19.6	735	2	UC5869
10	844.5	19.2	769	2	T51112
11	839	19.1	691	2	E70906
12	838.5	19.1	926	2	D83888
13	776.5	17.7	754	2	G90484
14	747	17.0	778	2	D72421
15	722	16.4	880	2	B45956
16	700	15.9	876	2	A45956
17	682.5	15.5	882	2	H82754
18	675.5	15.4	986	2	JC4825
19	642.5	14.6	765	2	AD0778
20	635.5	14.5	765	2	C64981
21	633	14.4	770	2	JC7728
22	628.5	14.3	765	2	C91006
23	626.5	14.3	743	2	D85850
24	625	14.2	743	2	E87386
25	609.5	13.9	764	2	F83431
26	609	13.9	806	2	D87627
27	593	13.5	825	1	GUHO
28	581	13.0	860	2	JC4939
29	572	13.0	763	2	D86136

30	566	12.9	763	2	JC4376	beta-glucosidase (
31	564	12.8	756	2	AD1422	beta-glucosidase h
32	547	12.5	740	2	H82807	beta-glucosidase x
33	546	12.4	727	2	AC0076	beta-glucosidase (
34	542.5	12.4	723	2	AG1662	beta-glucosidases
35	541.5	12.3	723	2	A11290	beta-glucosidases
36	518	11.8	776	2	T52390	beta-1,4-xylosidase
37	495.5	11.3	773	2	T49925	beta-xylosidase-11
38	491.5	11.2	792	2	T49983	beta-xylosidase-11
39	464.5	10.6	821	2	A49881	beta-glucosidase (
40	464.5	10.3	830	2	A44768	beta-glucosidase (
41	425	9.7	947	2	S08243	glucan 1,4-beta-gl
42	384	8.7	869	2	S24325	glucan 1,4-beta-gl
43	370.5	8.4	790	2	T49542	xylian 1,4-beta-xy
44	350.5	8.0	798	2	T00131	xylian 1,4-beta-xy
45	336	7.7	821	2	E87503	1,4-beta-D-glucan

## ALIGNMENTS

## RESULT 1

GLVK

beta-glucosidase (EC 3.2.1.21) precursor - yeast (*Kluyveromyces marxianus* var. *marxia*)

N:Alternate names: beta-D-glucoside glucosylhydrolase; cellobiase; gentiobiase

C:Species: *Kluyveromyces marxianus* var. *marxianus*, *Candida kefyr*

C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 28-May-1999

C:Accession: A29148

R:Raynal, A.; Gerbaud, C.; Francinques, M.C.; Guerin, M.

Curr. Genet. 12, 175-184, 1987

A:Title: Sequence and transcription of the beta-glucosidase gene of *Kluyveromyces fra*

A:Reference number: A29148; MUID:88210533; PMID:2835179

A:Accession: A29148

A:Molecule type: DNA

A:Residues: 1-845 <RAY>

A:Cross-references: GB:X05918; NID:92804; PIDN:CAA29353.1; PID:92805

C:Superfamily: beta-glucosidase

C:Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F:66,304,438,621/Binding site: carbohydrate (Asn) (covalent) status predicted

Query Match 46.7% Score 2052; DB 1; Length 845;

Best Local Similarity 47.2% Pred. No. 6.4e-132;

Matches 403; Conservative 152; Mismatches 269; Indels 30; Gaps 11;

QY	1	MADIDVEALIKRLTLAEKVDLAGIDFWHAKLPKHGVPSLRFTDGPNGVGRKFFNGVP	60
DB	1	MSKRDVEQLLSEINODEKISLSAVDEWHKTEILGIPAVRVSDDGNGIRGTRKFFDGPV	60
QY	61	AACPRCSTLSGTFNOLLEAGKMGKEALAKSAHYITLPTINMORSPLGGGFESIGE	120
DB	61	SGCPNCTGLASFDRDLLEAGKMGKEALAKSAHYITLPTINMORSPLGGGFESIFE	120
QY	121	DPLFAGLAALINGISTGVQATIKHFLCNDQEDRRRMVQSVTERALREIYALPQIA	180
DB	121	DPLFAGLAALINGISTGVQATIKHFLCNDQEDRRRMVQSVTERALREIYALPQIA	180
QY	181	VRDQPAFMTAYNGINGVSCSENPYLDGMLKREKMGDLGMSDWGTSSTFAVAGL	240
DB	181	VKHANPCIMTAYNVKNGVSCSENPYLDGMLKREKMGDLGMSDWGTSSTFAVAGL	240
QY	241	DLEMPGPRFEGELTKFNVSNGKPFIVH--IDQARVLOFV-----KCAASGVTEGPE	294
DB	241	DLEMPGPRFEGELTKFNVSNGKPFIVH--IDQARVLOFV-----KCAASGVTEGPE	294
QY	295	TTVNTPTETALLKRVNGEIVLLKNENNVLPSSKKKTLVGNPAQATYHGGGSAALR	354
DB	301	STSNNTKTSDDLREIADSIYLLKNNKNTLSKERRQYHYHGNNAKATSSGGASAMN	360
QY	355	AYIAYTPDGLSKOLEPPTSTVAGATYTP--PLIGECIIPDPGAP-----GMRVRFNE	407
DB	361	SYVAVSPYEGIVNKLKGEVDTVGAYSHKSGIGLAESLIDAAKPAEAENAGLIAKYSN	420
QY	408	PPGPRNRQIHIDELF---FTKTDMLIVDYHPRKAADT---WYADMEGTYYADECTYELG	460

```

Db 421 P--VERSEDEEPFVHTKYNRSNHLFDKHEKVDKPNPFVTLTGQYVQEDDYFS 478
OY 461 IYVCGAKAYVDOLVONATKOVPGDAFFGATREELGRINLVNGNTYKRIEFGAPT 520
Db 479 LOYVSSGLFYLDDELIDOKHNGEGRFCFGATKERRKTLTKLKGQYNNRVEGSGPT 538
OY 521 YTLKQDTYVPGHSLRVGCKVYIDQAEIERKSVLAKEDVOYICAGLNADMETEGADRA 580
Db 539 SLVGEF--GAGFGQAGVYKAIDDEDEIRNAELAAKHDAVLLIGNGEMETEGIDRE 595
OY 581 SMKLGVDLDOLIAVAANPNTVYVMTCTPEBMPILATPAVIOAWYGNETSADYV 640
Db 586 NMDLKRTEINLRAVLKANPNTVYVNOGTPEEPFMEENALVQAWYGNELNADYV 655
OY 641 VEGDNPBGKLSLSEPKRLQDNPAFLNFTAGRTLEGEDYVGRYREDFKDVNPPFG 700
Db 656 LYGDVVPNGKLSLSEPKRLQDNPAFLNFTAGRTLEGEDYVGRYREDFKDVNPPFG 715
OY 701 HGLSTTFAFSLVSHKDKLSVSKYKNTG-SVPGAQVQALVYKPLQAKINRPYKEL 759
Db 716 YGLSTTTELDIDSEKVIDDKIDISVDYKNTGDKFAGSEVQVYFSALN-SKVSRRPYKEL 774
OY 760 KGFARVELQPGETKAVTIEBOEKYVAAYFEEDRQMCVEKGYEVIYSDSAAADYALR 819
Db 775 KGFARVELQPGETKAVTIEBOEKYVAAYFEEDRQMCVEKGYEVIYSDSAAADYALR 831
OY 820 KGFVGETYVWMSGV 833
Db 832 KEFVKEDLYWKGL 845

```

## RESULT 2

```

A42292
beta-glucosidase (EC 3.2.1.21) - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 08-Oct-1999
C:Accession: A42292
R:Castle, L.A.; Smith, K.D.; Morris, R.O.
J. Bacteriol. 174, 1478-1486, 1992
A:Title: Cloning and sequencing of an Agrobacterium tumefaciens beta-glucosidase gene in
A:Reference number: A42292; MUID:92165721; PMID:1537792
A:Accession: A42292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1818 <CDS>
A:Cross-references: GB:M59852; NID:g142221; PIDN:AAA22082.1; PID:g142222
C:Superfamily: beta-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

```

```

Query Match 32.9%; Score 1445.5; DB 2; Length 818;
Best Local Similarity 40.5%; Pred. No. 1.7e-90;
Matches 334; Conservative 134; Mismatches 310; Indels 47; Gaps 19;

```

```

OY 6 VEAILKLTAEKVDLAGIDFMHTRKALPKHGVPISLFTDGPNGVR-GTRFENGVPAPCF 64
Db 2 IDDLIDKMTLEEDYSLISGADFVTVIAERLGTVPKIKYIDGPGANGGSLVGVASACF 61
OY 65 PCGTSLGISTNQTLLBAGKMKKEAIASAHVILPTINMORSPGLGCFESIGEDPPL 124
Db 62 PVAIALGATMDPELIERAGVALGQAKSGASVLLAPTVNIHRSNGNRECEYSEDPAL 121
OY 125 AGGAALINGISTGVQATIKHFLCNDDQEDRRMQSYLTERALREIYALPQIAVRS 184
Db 122 TAACAAYVINGVOSGVAATIKHFVANESEIERQNTSSVDERTLREIYLPPEEEVAKKA 181
OY 185 QPGAFMTAYNGINGVSCENPKYIDMLRKEMGMDLINSMDVGYSTTEAVVAGLDEM 244
Db 182 GVAWVSVYKNGKLTGTSEPMILLTVLREMGFDGVVSDMGFSHSTRETINAGIDLDM 241
OY 245 PGPPRRGETLKNVNSNGRPFTHVIDORAREVLYQVKKCAASGVTEGNETTVN--NTP 302
Db 242 PGPMRGRGKLVAAVREGVKAEVYASARILLILERV--GAEKAPDLAEHLADPE 298

```

```

OY 303 TAALLKRVNGEIVLLKNENNVLPSSKK--KTLVGPNAKQATYIGGSAALRAYAVT 360
Db 299 DRALIRQLGAEAGVALKND--GLVPLAKSSPDQIAVIGPNAASRVVGGSAIRAAITYVS 357
OY 361 PFDLSKQLTTPS--YTVGAYTT--VPTLIGBOCLTPDGAAPMRRVNEPPTGNRQH 416
Db 358 PLEBIRALSNANSLRRAVCCNNRRLIDVSGE--MTVEFKG--KGFESRPVHYEY 412
OY 417 IDELF--TKTDMHLVDYHPRADTWYADMEGTYTEADCTYELGLVYCGTAKAYVD 473
Db 413 KGEFVFDLPSGDLADLADF-----SARMATFVPQETGHIIGMTNAGLARLFFVG 463
OY 474 QIVVDAATKVPDPAFGSATREELGRINLVNGNTYKRIEFGSATYTLKQDTYVPGH 533
Db 464 ELVYDIDGWTGKNEFEGTANSQORAVTGAARRRVAYE--EARKASLDGINIC--- 518
OY 534 SLRVGCKVYIDQAEIERKSVLAKEDVOYICAGLNADMETEGADRAKMLPGVLDQIA 593
Db 519 ALRFGVEKPLGD--AGIAEAVETARKSDIYVLLVGRGEMDTGDLDPDRLPGRQEELE 577
OY 594 DVAAANPNTVYVMTCTPEBMPILATPAVIOAWYGNETSADYVFGDYNPSGKLSL 653
Db 578 AVAETPNVYVLTGTGPIEMPVLGRVAVLQWYVPGQELANADYVLFGEVPEAGRLPQ 637
OY 654 SFEKRLQDNPAFLN----FRTEAGRTLYGEDYVGYRYREDFKDVNPPFGHLSYTPFA 709
Db 638 TFPKALDMSATIDDPISITPGDGHVRAEGIFVGRHNDTEIEPLFPFGSLGTYTRFT 697
OY 710 FSNLSVS---HKDKLSVSLSKYKNTGVPGAQVQALVYKPLQAKINRPYKELGFAKV 765
Db 698 WGAPOLSGTEMGADG--LTATVDTNIGDRAGSDVQVYHS--PNARVERPPEKELBAFAL 755
OY 766 ELQPGETKAVTIEBOEKYVAAYFEEDRQMCVEKGYEVIYSDS 810
Db 756 KLPAGTATVAVLIAKIRDL-ATFDVAGRFADAGYELIYVASA 799

```

## RESULT 3

```

S04381
beta-glucosidase (EC 3.2.1.21) B - Clostridium thermocellum
N:Alternate names: Cellulobase B
C:Species: Clostridium thermocellum
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 24-Nov-1999
C:Accession: S04381
R:Graebnitz, F.; Ruecknagel, K.P.; Seiss, M.; Staudenbuer, W.L.
Mol. Gen. Genet. 217, 70-76, 1989
A:Title: Nucleotide sequence of the Clostridium thermocellum bglB gene encoding therm
A:Reference number: S04381; MUID:89364694; PMID:2505054
A:Cross-references: EMBL:X15644; NID:g40666; PIDN:CAA33665.1; PID:g40667
A:Note: part of this sequence, including the amino and carboxyl ends of the mature pr
A:Gene: bglB
C:Superfamily: beta-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:2-754/Product: beta-glucosidase #status experimental <NAT>

```

```

Query Match 26.1%; Score 1145; DB 2; Length 754;
Best Local Similarity 31.9%; Pred. No. 4.9e-70;
Matches 270; Conservative 132; Mismatches 249; Indels 196; Gaps 15;

```

```

OY 4 IDVEAILKLTAEKVDLAGIDFMHTRKALPKHGVPISLFTDGPNGVRGTRK-----FF 56
Db 3 VDIKITIOMTLERKAGICSGIDFMHTRKPVRLGIPSTIMTIDGPHGLKQKQEDAEIADIN 62
OY 57 NGVPAACPCTSLGISTNQTLLBAGKMKKEAIASAHVILPTINMORSPGLGCFE 116
Db 63 NSVPATCPPSAAGLASMDELVERVGAALGEEQANVSLILGPGANIKRSPICGNFE 122
OY 117 SIGEDPFLAGAAALLRGISTGVQATIKHFLCNDDQEDRRMQSYLTERALREIYALP 176

```

```

Db 123 YEPEDYLSSELASHIKVQSGVACLTKEFAANNOEHRRMTDTIYDERTLEIFEAS 182
Qy 177 FOIAVRSQAPAFMTATNGINGVSCSENPATLDGMLKREKMGDGLMSDWITGISTEAV 236
Db 183 FENAVKARPMVAVVCAYKKNLNGEYCSERNYLLTEYLKEMMHDEFFVSDWCAVNDVRSGL 242
Qy 237 VAGDLEMPGPPRFGEFTLKFNVSNGKPFHIVIDORAREVLQFVKCAASGVTEGNETT 296
Db 243 DAGDLEMPTRSHGITDKTIVARVSKGISENINLRARERILAKITM-----ALEKKENA 297
Qy 297 VNNPETAALLRKVNGNEGIYLLKNENNYLPLSKKKTLIVGPNKQATYHGGGSAALRAY 356
Db 298 QYEDAHRLARQAAAESEVLLKNEDDYLPLKSGTIALIGAFVKRPYQSGSSHI--- 354
Qy 357 YAVTFDLSKOLETPPYTYGATTVPTPIIGEOCLPRDGAQRKRWVFNPEPTPNRQH 416
Db 355 -----TPTR-- 358
Qy 417 IDELFETKTDHLYDYHPRKADTWYADMEGTATADECTYELGLVYCGFAKAVVDOLV 476
Db 359 LDDI-----YEIRK-----KAGAD----- 372
Qy 477 VDNATKQVPGDAFFGSATREETGRINLVKGTYFKLIEFGSAPTYTLKGTIVPGHSLR 536
Db 373 -----KVNLYVSEGYR-----LENDGI----- 389
Qy 537 VGGCKVIDDOAEIEKSVALKHEHOVITICAGLNDWETEGADRASMLPGVLDLIDVA 596
Db 390 -----DEELINEAKKAASSSDVAVFAGLPDEYESEGDRTIMSTPENONLLIAYVA 441
Qy 597 AANPTVVVMOTGPPEEMFPLDAPPAVIOAWYNGENETN-----SIADVYFGDYNSGK 650
Db 442 EVQSIIVVVLNGSVYEMHWIDKVSYLEAVLGGALGGRMRMCYSKSYV-----GK 494
Qy 651 LSLSPKRLQNPAPLNFRTAAGRTLYGEDVYVGRYRYEFPADKQVNPFPFGHLSYTPFAF 710
Db 495 LAETFPVLSHNPSTYLNPGEDDREYKEGELFVGRYDYDTGIPLEPFPFGHLSYTKFEY 554
Qy 711 SNLSYSHD-----GKLSYLSYKNTGSPVQAOVLVYKPLQAKIRPVEKLGFAKVE 766
Db 555 SDISYDKKVDSDNITINSVKKNVKGAKGKELVOLYKVDKKS--VRPEEKLGFEEVF 613
Qy 767 LQPGETKAVTIEOEKYAAAFDEBDQMCVEKGDY-----EIVSDSSAAKDGVA 817
Db 614 LNPGEKIVTF-TLDKRAFAYYNOIKOMHVESEGFLLIGRSSRDIVLKESVAVNSTVK 672
Qy 818 LRGFYV 824
Db 673 IRKRFYV 679

```

## RESULT 4

beta-glucosidase (EC 3.2.1.21) - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: C72428  
 R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A:Reference number: A72200; MUID:99287316; PMID:10360571  
 A:Accession: C72428  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-721 <ARN>  
 A:Cross-references: GB:AE001690; GB:AE000512; NID:94980496; PIDN:AAD35119.1; PID:9498050  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0025  
 C:Superfamily: beta-glucosidase

C:Keywords: glycosidase; hydrolase; polysaccharide degradation;

Query Match 21.8%; Score 959; DB 2; Length 721;  
 Best Local Similarity 30.4%; Pred. No. 2.3e-57;  
 Matches 262; Conservative 130; Mismatches 280; Indels 190; Gaps 26;

```

Qy 6 VEAIIKLTLAEKVVLDLAGIDF-----WHTKALPKHGVPSLRPTDGPNGV 50
Db 4 IDEILSQLTTEERKVLVGVGLPGFNGPNHSRVAGAAGETHVPRLGIDPAFVADGPAGI 63
Qy 51 R----GTFKPNVPAACFPGCGTSLGTFNOTLLEEGKMMKKAIAKSAHVILGPTINQ 106
Db 64 RIPTRENDENITYTTAFPEIMASTWMDLLEEGKAMGEVREYGVYDLAPAMNH 123
Qy 107 RSPLGGRFESIEDPPLAGLGAALIRGIOSGYOATIKHFLCNDQEDRRMVOSTIVE 166
Db 124 RNPGLGRNFEEYSEDPLSEMASAYKGVQSGVACLTKEFAANNOEHRRMTDTIYDE 183
Qy 167 RALREIYALPFOIAVRSQAPAFMTATNGINGVSCSENPATLDGMLKREKMGDGLMSDM 226
Db 184 RALREIYALPFOIAVRSQAPAFMTATNGINGVSCSENPATLDGMLKREKMGDGLMSDM 243
Qy 227 YGYSTTEAVVAGLDLEMPGP-----RFGEETLKFNVSNGKPFHIVIDORAREVLQ 279
Db 244 YAGDNPVEOLKAGNDMIMPCKAYOVNTERDELEETLEALKEGKLSSEVLEDECVNRILK 302
Qy 280 VKKCAASGVTEGPE---TVNNTPTPT---AALLRKVNGEYLLKNNENNYLPLSKKKKT 333
Db 303 -----LVNAPSRKGYRYSKPKDSEHAEVAYENAGAGVLLER--NGVLPFDENTHY 353
Qy 334 LIVGPNKQATYHGGGSAALRAYAVATPFDLSKOLETPPYTYGATTVPTPIIGEOCLT 393
Db 354 AVFGTQIEIKGTGSGDTHPRYTISILEGIKE----- 387
Qy 394 PDGAPGRMREVFENPEPTPNRQHIDELFTKTMHLYDYHPRKADTWYADMEGTATDE 453
Db 388 -----RNKPFEEELASTYEYIKR-----MRETEYKPR-TDSW----- 420
Qy 454 DCTYELGLVYCGFAKAVVDDOLVVDNATKQVPGDAFFGSATREETG--RINLVKGTYFK 511
Db 421 -----GTVI-----KRLPENPLSEKEIKR-----AAKNDVAVVVISRISGSGYDR 462
Qy 512 KIEFGSAPTYTLKGTIVPGHSLRVGGCKVIDDOAEIEKSVALKHEHOVITICAGLANA 570
Db 463 K-----PVKGDYFLS-----DDELELKITV--SKFEHQ----- 489
Qy 571 DMETEGADRASMLPGVLDLIDVAANPNYVVMOTGPPEEM-FMLDAPPAVIOAWYG 629
Db 490 -----GKVVVLLNIGSPIVAVSMRDVLVDGILLVWQA 521
Qy 630 GNETGSIADVYFGDYNSGKLSLSPKRLQDNPAPL---NFTTEAGRTLYGDDVYVGR 686
Db 522 GQEMGRIVADVLGKINPSKLTPTTFPKQSDVPSMTFPEPBDNQRVYEEEDIVGVR 581
Qy 687 YVEFADKQVNPFPFGHLSYTPFAFSLVSHKDGKLSLSYKNTGSPVQAOVLVYK 746
Db 582 YDTFGVEPAYEREGYSLYKFEYKDLKIAIDETLRVSTYITNTDRAKGEVSOYVYIK- 640
Qy 747 LQAKIRPVEKLGFAKVE-LQPGETKAVTIEOEKYAAAFDEBDQMCVEKGDY 805
Db 641 APKGTIDKPPQELKAFKFKYLLNPGSEETSLPIPLDASFDGKE---WVSEGEIEVR 697
Qy 806 VSDSSAAKDGVALRGFTV-GE 826
Db 698 VGASSR---DIRLRDIFIVEGE 716

```

## RESULT 5

T35785  
 probable beta-glucosidase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
 C:Accession: T35785

Query Match 21.8%; Score 957; DB 2; Length 859;  
Best Local Similarity 29.7%; Pred. No. 4.2e-57;  
Matches 266; Conservative 131; Mismatches 354; Indels 146; Gaps 22.

Qy 754 RPVKELGFAKVELDPGETKATIEEQEKVAAYFDERDQWCYEKGDYEVISDSS 810  
 || : | : | : |||| : : : | | | : : | :  
 Db 787 RPARWILGYTAVRARPGETVATVRPAR-ALRHWSVAEHAWRTTEAGPCRYLAGRSA 842

beta-D-glucosidase [Imported] - *Caulobacter crescentus*

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-767 <STO>  
A:Cross-references: GB:AE005673; NID:g13423178; PIDN:AKK2732.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CCL1756  
A:Superfamily: beta-glucosidase

Query Match	21.6%;	Score 946.5;	DB 2,	Length 762;
Best Local Similarity	30.5%;	Pred. No. 1.8e-56;		
Matches 257;	Conservative 115;	Mismatches 273;	Indels 199;	Gaps 20,

QY	4	IDEALTKLTKLTAKEVNDLLACI-----DPMHTALKRHKVPSLFTDPPNG	49
Db	55	VRRDALVAKMTIDEXITYLHLEFPMAKRAPDTMISACVYGVRLMPTLRESOASLG	114
QY	50	VRG-TKEFNGVPACFPCGSLGSTENOTLLEBAGKMKKALIAKSAHVILGPTINMORS	108
Db	115	VANVOQRKGDVAZTLPSGLMASTFEPRKLAFGGAMIGAEARAFTFNVLGVNLTLD	174
QY	109	PLGGREFEIGEDPLFAGLAAALIRGISTOGVATIKHFLCNDDSDRRMVOSTLTERA	168
Db	175	PMAGRFEEILGSDPLLAGEMVAEDQIKVOYNKTVSTIKKFLPMAOETGRHYMDQIDBAD	234
QY	169	LREIYALPFOIAYVRDSOPAGFMTAYNGINSCSENPKYLDGMLRKEWGDGLIMSDWG	228
Db	235	LRESLDLAFQAIERKSNPASYACVAKYKNGVNDMACENDPLFNKYLKRDNNYEGWYMSDGA	294
QY	229	TYSTTAAYVAGDLEMPERP-----FKGELTKNNVNGNCRPLFIHIDQARAVLOFPYKCA	284
Db	295	VHSTVAAALAGLD-QOSSQELDTQIFRGEDLKAAYKAGEVSOARVDNRVRLILGV---I	350
QY	285	ASGVTEGNETEVN--NNPETAALLRKVNEGIVLLKKNENVLPUSLKR-KRTLLVGNPAK	341
Db	351	SSGLMDNPLPTSAGQIDYDAIAKAAQVYARSGVSLKLNDRGLPLPAKSAKIVYLGAHAD	410
QY	342	QATYHGGSSAALRAYAVTPDGLSKOLETPSYTVGAAVTPVPLLEGCCLTPDGABGMR	401
Db	411	VGVISGGSSQYR-----SVGAPVEIPLKGEBAAS-----	441
QY	402	WRWFNPNPPTPRQHIDELFEPTKTDMLLVYHYPKAADYVDMGTYTAEDDC-YELGL	461
Db	442	-----FVRV-----TWBA-----	449
QY	462	VVCGTAKAVDDOLVVDNATKOVPGDAFEGSATREETGRINLVKNGTYKFIIEGSPATY	521
Db	450	-----SSPLQ	454
QY	522	TLKGDTIVPGHSLRGCKYIDDAETIEKSVALAKENDQVITCGLNADMETGADRAS	561
Db	455	AIKA-----ANPSAEV---TYVGSKDPV-AAAAAAMDADVAIVFAM---HHQTEADAPS	502
QY	582	MKIPGVLDLADYAAANPNVVMQGEPEPMPLDPAIYATQWAGNENGSISIDV	641
Db	503	IALPENODLLEANSAAKNAVVULETGVGLPMLDKKAVGLQWYTCGORGGAIRLL	563

QY 642 FGDYNSGKLSSEPKRLQDNP-----AFINFRTEAGR-----TLGEDVY 682  
DB 563 FGEVNSGRLAMTFPKSEDOAPRASPAPFAEDQAIIDARRAGOKAPINGFPVRYEGNA 622  
QY 683 VGYRYEFADKDVNPFPGHGLSYTFEAFNSLSYSHKDGKLSVLSYKNTGSVGAQVLAOL 742  
DB 623 VGYRMFAOEKRRRLPYFPVGLSYTSFGYKMLKVEDGDG-LKVSFDVTNTGKVAAGADTPOL 681  
QY 743 YVPELQAKINRPVKELGFAKVELOPGETKAVTIEOEKYAAVYDEREDOMCVKKGXY 802  
DB 682 YVTSQG-----RKAMRLAGFQKVDLAPGETKRVTLNVEBRILLADY-DTAKPGMTTAAQY 736  
QY 803 EVIY 806  
DB 737 PLTY 740

RESULT 7  
AF0341  
probable beta-glucosidase (EC 3.2.1.21) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AF0341  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.  
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0341  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-793 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CA93037.1; PID:q15980775; GSPDB:GN00175  
C:Genetics:  
A:Gene: bg1B  
C:Keywords: glycosidase; hydrolase

Query Match 21.4%; Score 939.5; DB 2; Length 793;  
Best Local Similarity 28.7%; Pred. No. 5.8e-56;  
Matches 242; Conservative 125; Mismatches 266; Indels 211; Gaps 16;

QY 6 VEAIIKLTIAEKVDLAGIDF-----WHTKALPKHGVPSLFTDGPNGVR--- 51  
DB 31 IPSVIAMTLEKKAFAVSGTGMNAEKVAGAAGSTLAIPLGIPIYFADGPGVRLGAG 90  
QY 52 ---GTRFNGVPAACFPGCTSLGTFNQTLLERAGKMKKEALAKSAHVILGFTINQRS 108  
DB 91 PTGGEKRF---ATGFVSVAMATWDAIKRYGGAIGDEAKOYGVDLILGPAINIORH 146  
QY 109 PLGGRGESIGEDPEFLAGLAALIRIGIOSTVOATIKHFLCNDQEDRRMMVSIYTERA 168  
DB 147 PLKRNNEFYTEDPLNAINASTIDGMOTEGVAILKHAANNOETRRQIVIEITSDRA 206  
QY 169 LREIYALPFOIAYDSOPGAFMTAYNGINGVSCSENPKYLDGMLRKEMGMDLIMSVMY 228  
DB 207 LHEIYFGFEYAMKKAQPMWMSYPSINTPSSQNPWLKDVLRQOMKRDGVMDSWYG 266  
QY 229 TYSTTEAVAGLDEMG-----PPRRGETLAKNNVSNKGFPIHYIDQAREVL 277  
DB 267 VADVPASALKGNDLMPGGRTPDSDLFTPNTDPKDVLAALKSGELTODODIENINIL 326  
QY 278 QFYKCAASGVTENGPEPTVNNPETALLRKVNEGIVLLKNNNNVLPISKKKTLIYG 337  
DB 327 NVVTK---THREFKGDAAAMEPEVDHSAALQAIADSMVLLKNDARHTLPKAAFRVAAG 383  
QY 338 PNAQOATYHGGGSAALRAYAVTFPDLISKOLETPPSYTGAVYTVPIIGEOCLTPDGA 397  
DB 384 QNIDNFPYTGGSAAEVN---IDP-----KRLVTLPGSLNAGIALIRQIGDQAL----- 429  
QY 398 PGMRRVFNPEPPTPNQOHIDELFTYKTDMLHLDYHPRKADTWYADMEGTYADEDCITY 457

DB 430 -----DEPVNDP----- 436  
QY 458 ELGLVCGTAKAYVDQDLVNDATKQVPGDAFFGSGATREETGRJNLVGNTRYKKEIERGS 517  
DB 437 -----ALTRA----- 441  
QY 518 APTYTLKDDTIVPGHSLRGCKVYIDQAEIENSVALKAEHDOYITCAGLNADMEEGA 577  
DB 442 -----AQOSDIALISIGRSS---TEGA 460  
QY 578 DRASKKLGVDLQDLADYAAA-----NPTVVMQGTPEEK-PWIDATPVIQMYGNE 632  
DB 461 DRYSAMHPDEYAMIKKYASAFHQNKRVYLLNIGSIEMTSQVADALLTWMQPEEQ 520  
QY 633 TGNSTADYVFGDNPSPKLSSEPKRLQDNPAFLNFRTEAGRTTLGEDVYGYRYEPAD 692  
DB 521 AGNAVADIIIGKVNPSGKLPITFPKRLDPSFGYPPGNATVYIGDEIYGYRFDKR 580  
QY 693 KDVPNPFHGLSYTFEAFNSVSVS-----HKDKLSVLSYKNTGSVGAQVLAOLYKP 746  
DB 581 IAPWPFYGLSYSTVYNGKISPEKRVFNIDTENSIEVSIPIRNTSGIDTREVOLYVHD 640  
QY 747 LOAKINRPVKELGFAKVELOPGETKAVTIEOEKYAAVAFDEREDOMCVKKGXYEY 806  
DB 641 -NASRLDREOELKAFKESVLSAGEKRYAF-KIDKRAFSTYDEKNNMVEPLFTLRI 698  
QY 807 SDSS 810  
DB 699 GRSS 702

RESULT 8  
D87369  
beta-D-glucosidase [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: D87369  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: D87369  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-758 <STO>  
A:Cross-references: GB:AE005673; NID:q1342248; PIDN:AKK22952.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0968  
C:Superfamily: beta-glucosidase

Query Match 19.9%; Score 874; DB 2; Length 758;  
Best Local Similarity 29.7%; Pred. No. 1.6e-51;  
Matches 257; Conservative 102; Mismatches 286; Indels 220; Gaps 23;

QY 2 ADIDVEAIIKLTIAEKVDLAGI-----DFWHTKALPKHGVPSLRF 43  
DB 49 AOKRADVLVAKMTNDEKFTLIRYEGADMKPYTRIPESLPSAGYIAGVPRHIGIPAOFO 108  
QY 44 TDGPNQVR---GTRFNGVPAACFPGCTSLGTFNQTLLERAGKMKKEALAKSAHVILG 100  
DB 109 TDAGVAVATQGSAKERR--ERTALPBGAMTATWNEFLAFQGGAMIGKARDSGENVOLA 166  
QY 101 PTINMQRSPUGRGSEGEDPEFLAGLAALIRIGIOSTVOATIKHFLCNDQEDRRMY 160  
DB 167 GGVNLVREPRNGRNFYEGEDPLLAGVMAADIRIGIOSNNIISTIKHYLANOEGRGFV 226  
QY 161 OSIATERALREIYALPFOIAYVDSOPGAFMTAYNGINGVSCSENPKYLDGMLRKEMGMD 220  
DB 227 SSNIIDTARMSDILAFOAIBOSDPHSYWCAYNRVNSYACEPDLNLTVLTKDMGYKG 286  
QY 221 LIMSMDYGYSTTEAVVAGLDEMG-----PPRRGETLAKNNVSNKGFPIHYIDQAREV 276

[illegible]

2 ADIDEALIKKTTLEKVDLLAGID-----FMHKKRALPHGVSLRFED 45  
 37 ADARARQVLAASLEDEKMSLLFVSVDGGGNGSVAPRGGLSAALYLRARQSGSLPLQLQISD 96  
 46 GENGVRGTKFF--NGVPAACFPCGTSLSGTFNQTLLLEAGKMGKEALAKSAWILGPTI 103  
 97 AGLGRNPAHRIHRR--EAVSLPSGOSTASTWMDMARQAGVIGREAWOSGFENILLGGA 155  
 104 NKQSRPLGGRGREGESIGEDPFLAGLGAALIRIOSTGQVATIKHFLCNDQEDRRMAYOSI 163  
 156 DLTRPDGRGRNVEYGEORPLQTRMVGSTIAVQSOAHYSTLKAHAMNDELTSRMTMSAD 215  
 164 VTERALREIYALPQIAVNDSQGAFMTAYANGINCSVENPKYLDGMKLRKEMGIDGL 223  
 216 IDPVAMRESDDLGFETALFETGHPGAVKMSYKNVNDLYACENYLLKTKLKQDMHIFGVYM 275  
 224 SDWYGTSTSTTEAVAGLDLEMPG-----PRRGFTLKFVNSNGK--PFIHVLDQARREVL 277  
 276 SDMGATHTSSARAAALGLDQESAGDHTDARPYFR--TLIADVYAGVAPPEARINDMAER--- 331  
 278 QPVKKCAASGVLEN-----GPEYTVNNTPEYTAALLKKNVNEGIVLLKNNVPLSKKKKT 333  
 332--VYRALFAAGLVDHNAQSKRPLDVTDT-----LVAAKDEEGAVLLRNGNMLPLSPARI 386  
 334 ILVGNNAQATFHGGGSAALRAYAYAVTPFDLSKQLETPPSYTGAVYTPVPLICEQCLT 393  
 367 AVYGHADAGVYSSGGS-----SQVDPDGE----- 412  
 394 PDGAGCMKRWVFNPPGTPRNQHIIDELEFTYKTMHLDVYIYHPRKADTWYADMEGYTIADE 453  
 413----- 412  
 454 DCTYELGLVCGTAKAYVDDQLVVDNATKOYRPGDAFPGSATREERGRINTLVKGNTRYFKI 513  
 413-----AVKGPGRK-----KEMPGDEVY----- 428  
 514 EFGSAPTYTLKGDTLTPGHGSLRWGGCKVIDDOAEIEKSVALLAKEHDPVILICAGLNADME 573  
 429--FPSPSLAKMAQAE--APG---ARI---TYDPGTSIASAVALAARADVVVYVA---TQFT 475  
 574 TEGADRASKMKLPQVLDLIADVAANPNTVVYVMQGTPEEMRPYLDATPAVYQAMYGNET 633  
 476 FEGMDAPSMHLDNDNADALITVAANPRTVVYMETGDEVILPMMNSVAVGLEAVFPGGSG 535  
 634 GNSIDVVEGDPNPAGKSLSPKPR-----LDQMPAFINFTTEAGRTLYGEDV 681  
 536 GPATIRLLFGKVAAPBSGHLTMTTPPOAESQLAHPDLIAGVADAVFEKQFTTDO--ELYIDEGS 594  
 682 YVGKRYEYFADKDVNPPGHSILSTTFAPSNLSVSHKDGKLSVLSYKNTGSPQAOYVQ 741  
 595 DVGWKFDFRNHFKPLPYRGYGLTYTTFSTDGLKTERHGQVYATAFNVNHTGTBRAGVDVQ 654  
 742 LYVAPLQAAKINRPYKELKGFAYKVELDQGETKATYIEQEKYVAAYFPEEDNDQKCYEKG 801  
 655 YVY-----GLPDGGAARRLAIQMORISLAPGESROYSV--OLEPRLLAHFQKHDRMSVPSGT 708  
 802 YEVIYSDSSAAKD 814  
 709 FRVWL--ASCATD 719  
 RESULT 10  
 T51112  
 beta-glucosidase (EC 3.2.1.21) oler [validated] - Streptomyces antibioticus (ATCC 118  
 C:Species: Streptomyces antibioticus  
 A:Variety: ATCC 11891  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
 R:Quinos, L.M.: Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J.A.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z25300  
 A:Accession: T51112



A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-769 <QUIT>  
A:Cross-references: EMBL:AF055579; PIDDN:AAIC12650.1  
A:Experimental source: ATCC 11891  
C:Genetics:  
A:Gene: oler  
C:Function:  
A:Description: extracellular reactivation of oleandomycin; converts glycosylated oleand  
C:Keywords: glycosidase; hydrolase

Query Match	Similarity	19.2%	Score 844.5	DB 2	Length 769
Best Local Similarity	10.0%	Pred. No. 1.7e+49			
Matches 253	Conservative 126	Mismatches 332	Indels 133	Gaps 26	
13	LTAAKVDLLACIDFW--HTK-----	ALPKHGVSLRFTDQNGVGRGKFFNGCAPA	61		
1	MTLEELKISFV----	HMSYTTSDESAQKVVLDGVPRLIGLIPEMRAIDAGGATIIH-----	PS 51		
62	ACFPGCTSGSTFNGNLTLEAGKMGKKEALIASAHVILPTTINMORSPLGGRFESIGED	121			
52	LALPAPVALASTFDDGLASVYAGVIGREBRAGQDVFAFPMVNSIVPAAGRFEETFSED	111			
122	PELAGIAAALIRGIQSTGVQATIKHFLCNDQEDRRMAVOSIYTEALKEITVLPQIAY	181			
112	PLVTSMAAEIKIGIOSGLIATKHYAANNOKNFFSVANVNDQOTLRERELPGEASV	171			
182	RDSQPAEMTAVNGINGVCSENPKLTDGLKREKEMCDGLIMSQVGYSTTEVAVAGLD	241			
172	A-AGTSVMCAIKKAVGQAPACSGDELLINKVLKQMKFRGWNISDWLQTS-TDALIRGLD	229			
242	LEW-----	PGPP--RFRGETLKEVNSGKRPVIAHIDQARAVYLFVKKCAASGV	288		
230	QELGIELDHPHAPGERPIPGKPFEGDPLKTAIRGRIPIESALDEAVVIRIYSOMARFL--L	287			
289	TENGPEPTYNNTPETAALLRKVNEGIVILKNNNVPLSKKKKT--LIYGPNAKQATYH	346			
288	DEPPRPARPADLAGLKVAQVAEDGVALLRNBSGAILPLTTERAADIAYIGPRAPKVT	347			
347	GGGSAALRAYAATPEDGSKOLETPPSYTVGAITYVPIIGEOCL--TPDGAPEGMR	403			
348	GLSS-----	YIYP--DGASAPLDITIRE-RAGAGSTVRYSTGEETVGVVPVQSAPIRLPR	399		
404	VNEPPCTPRKHIDELFFTKTDMHLVDYHHKRAADWYADMSTYTADECTYELGLVY	463			
400	-----	PSG-----EYFAGGGGVLD-----	GLTVPVYTGSRIT-----	428	
464	CGTAKAVVDQVLVDNATKQVPGCAFPGSATREETGRINLYKNTYFKIEFGSAPTYL	523			
429	-----	AARQGGNAIYELDQGEPRGRRPVYG-----	DVSSRPM--	462	
524	KGDITVPGHSLRVGCKVIDD-----	QAEIFKSYALKKEHQVLIICGLN	569		
463	---RLAAGTHLRLRTGAALKAKSPMTFELTWVTPQAADEALIDRAVSIANTATAVFA--	516			
570	ADMETEADRAASMTLPGVLDLIADVAANPNYVVAQOTGPEEMPMIDATAFVIAQWYG	629			
517	YDDGSEEGDRSLSLPRQODLISAVAAVNRVTYVINTGSSILMPMLRKTAVALTWMP	576			
630	GNETGNSIADVVEGDYDNPSSGLTJSFPRKLQDNPAFLNFKTEAG--RTLYGEDEVYCYXR	686			
577	GQAQAEATTTALLFDADDPGRGLTYTFPADEGOTFFAGDARYPGVDDDLDTSEGIYSGR	636			
687	YEEFADQVNPPEFGHSGSYTYFAFSNLSV--SHKQSGKLSVLSVKNTSQSVPAQVAYQALYVK	745			
637	WEQOQVQPLSPFGHGLCYTISFDYRDLKVTAATAGGLDVCFTLRNTGRTGCKEYVQVYVG	696			
746	PLQAAKINRPVKELKGFAKVELDQGETAAVNIIEDEQKVAAYFDEBERQWCEKQDYVI	805			
697	PSPHRVAVQAOKRAALAAVKGVELRPGESRRTL--HYERRALONWDSGAHTWYTGPG--ROYM	754			
806	VSDS 809				

Db 755 VGFS 758

RESULT 11

E70906

probable beta-glucosidase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: E70906

C:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulstron, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9654220

A:Accession: E70906

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1691 <COL>

A:Cross-references: GB:297050; GB:AL13456; NID:g3256008; PION:CAB09737.1; PID:g221355

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: bglI

C:Superfamily: beta-glucosidase

Query Match	19.1%;	Score 839;	DB 2;	Length 691;
Best Local Similarity	29.7%;	Pred. No. 3.5e-49;		
Matches	254;	Conservative 103;	Mismatches 283;	Indels 216; Gaps 25;

  

QY	13	LTLAEKVDLLAGI----	DEFW-----	HTKALPKHCVPSLRFTDGGNGVARG	52
DB	1	MDDEDFSLVGTIGTASDLMPVDERIPG	GVPMGAYVPGIRLGVPAALMSDAGLGYTN	60	
QY	53	TTFNNVPAACPPCGTSLGSTEFGNTLLE	AGKMGKGEAKSAHVILGPTIIMQSRPLGG	112	
DB	61	PGRPDDITATPAGATLAAFNRPVLA	RSSGAIGREARSRGFNQIAGAILANDPRNG	120	
QY	113	RGFESIGEDPFLAGLGAALIRGIQSTG	VOATIKHFLCNDQEDRRMVOСИTERALREI	172	
DB	121	RNFYEISEDPILSATWMAASIIGIQOG	VIATTKHFSLNCNETNRHMLDAVIDPDAHRS	180	
QY	173	YALPQIAVRDSPGAFMTATYNGIN	GVSCENPKYLDGLRKREMGDGLJMSDMYGTST	232	
DB	181	DLTAEFIVIERSPGAVMAAYNKNG	VDYAAGNDHLNDLVLKGMGIRGVAWSDMGSTPS-	239	
QY	233	TEAVVAGLDLMEWG-----	PPRFRTGELLTFNVSNGKPFITHVIDQAREVLOFV	280	
DB	240	WECALAGLDQECAGQIDAIVLMQSE	AFTRDRLRAAYVADGNLPKGR-----	291	
QY	281	KKCAASGVT--ENGCPETVYNNP	ETALALRKVGNBGIYVLKRNNNYLPISKKK--	RTLIIV	336
DB	292	RSMFVAVIDHWKPAAPADMANH	NEIAA--QMARQISIVLQN--RGLPLAPESAGIAI	VI	347
QY	337	GPNAKQATYHGGSSAALRAVAY	TPFDGLSKQLETPSPSYVGAVTTPPIIGBQCLTP	DDG	396
DB	348	GGTAHLGVPAGYSSAV-----	TPP--GGYAGYIPI-----	G	377
QY	397	APGMRWVENBPPTPNKQIH	LEFTTKDMLVDYIYHKAADTVADMGCTTAD	EDCT	456
DB	378	GGSL-----			381
QY	457	YELGLVGCYAKAVYDDQLV	VDNATKQVPGDAFFGASATREETRGINLVK	GNTRYFKRIEFG	516
DB	382	-----AAGLRNLILPSS	FLSELKRRLP-----	NAQFERPDG	413
QY	517	SAPTYTLKGDITVPGHSL	RNVGGCKVIDQAEILEKSYALAKKHEDV	YIICGLMAADMETEG	576
DB	414	INPA-----	EAVLAARRADIAIVFA--IIR--EGEG	440	
QY	577	ADRSAMTLPGVLDQILAD	VAAANPTVVVQGTPEEMPLIDATPA	VIQAWMGNETGNS	636
DB	441	FDSADLSIPWGQDILAA	VASANANTVYLETGHPYIMPRDVSNA	IMQAWMGQAGQGA	500

QY 637 IADVVEGDVNPSSKLSLSPKRL-----ODNPAFLNRETAGRTL-YGEDVYVGYRYERA 691  
 Db 501 VAEIVAGVNPSSRLPITTEPVDGCPRRSQPPELGNPMGTSTIHTIEADVGYRFEAST 560  
 QY 692 DKRVNPPFHGSLSTYTFASNLVS--HKDGKLSVLSVKNMGSPGACVADLYVKKPLDA 749  
 Db 561 NQPMFAFGHSLSTYTFEYRDLVLTGGH--TVHASFSVTNTGDRSGADVPQLYM-----I 613  
 QY 750 AKINRVEKELGFAKVELQPGETKATIEDEKRYAAAYDEDERDQMCVEKGYEYVSDS 809  
 Db 614 AAFGEERLKLGFERYELPEPGOTRRVRIADRLRLARY-DGEARSRIEPPGGYTAAGAS 672  
 QY 810 S-----AARDGYALRG 820  
 Db 673 AVALKLAARKVLAGRG 688

## RESULT 12

glucan 1,4-beta-glucosidase BH1908 [imported] - Bacillus halodurans (strain C-125).  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: D83888  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: D83888  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-926 <STO>  
 A:Cross-references: GB:AF001513; GB:BA000004; MID:910174345; PIDN:BA005627.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH1908

Query Match 19.1%; Score 838.5; DB 2; Length 926;  
 Best Local Similarity 27.3%; Pred. No. 5.9e-49;  
 Matches 241; Conservative 163; Mismatches 311; Indels 169; Gaps 32;

QY 3 DIDEVALIKLTLAEKVDLAGIDFWHTKALPKHGVSRLRTPDGPNGVGRGTFNGVPA-62  
 Db 18 DERESIVSELTDEKIRLMC---QVQEAIVRLGIG--KKHGTBGAHVAMIG--KAT 69  
 QY 63 CFPCTSLGTFNQTLEAEKMGKKAIA-----KSAHVILGPTINNQRSPFGGRF 115  
 Db 70 VFPPNIGLGTWNEELMKKIDVIAEARVYVOENPRONGLTWAPVDMERDPRMGRT 129  
 QY 116 ESIGEDPFLAGLAALIRGIGS-----TGVOATIKHFLCNDODRDMAMQSIYTERALR 170  
 Db 130 EAYGEDPFLITLTELKMGQSHFYLKATILKHLFLANNDEARTTECSVSIDPRNLR 189  
 QY 171 EYVALPOIAVRDSQCAFMTAYNGINGSCSENPXYLDMRLKEMWMDLIMSD----- 225  
 Db 190 EYULKAEPDYKKEGHVOCVMTAYNAVNGTLCNNMPD--VNLILKEEMGLQGFVSDACDV 248  
 QY 226 -----WGYTSTTEA--VVAGLDLEMPRPFRPGRGTLEFNVSNGKPFTH----- 267  
 Db 249 GSVUDHOFVESIYAAVALSTKNGID-----NYTDEKIIHRAIGDALEQ 292  
 QY 268 -----VIDORAREVLOFVKCAASGVTEGNETTVNN---TPEDAAILRKVNGGIYL 317  
 Db 293 GLTSEEDLDQAKRVFRVRIERLEFDEKNNPATYAKKELCAPHEISELALQAKAIVL 352  
 QY 318 LKNNNNVLPSSKKKTLIVPNKQ--ATYHGGSAAALRAYAVTFPDGLSKOLETPPSV 375  
 Db 353 LKND-DLPLSPKRYAVIPLANEIYTDWYSGTTP-----YIISPLEGIMQKANPVLX 406  
 QY 376 TVGAYTVVPIILGQCITPDGACGMRRVNEPGRPNRCHIDELFFYTDMLVDVYHP 435  
 Db 407 A-----DGSRIIRLRAMSGNGIVAEGSSOL--TCHDGGGTFTH 446

QY 436 K---AAPTWYADMESTY--ADECTYELGLVCGTAKAYVDOLVVDNATKQVPGDAF- 489  
 Db 447 DMGCSFELQSOQSKYTTLNDEGVLANALEVKGW--FVKELAMEENDHEVTLHTWD 503  
 QY 490 FGSATREETGRLINLVKNGTYFKKIEFGSAPTYTLKGTIVGCHGSLRVGGCKVDDQAE- 548  
 Db 504 RQVYKRDESERLS-VNGB-----NDCSLVKDLYER 532  
 QY 549 -IEKSVLAKEHDOVILICAG---LNADWETEGADRSMKLPGLVDLIADVAANENTV 603  
 Db 533 GIDDAVEAKQADAVAVFVGNQPYING---KECVDRBDLAPPAQROLIEEVYKVNNTV 589  
 QY 604 VVMQGTPEEMPTL-DATPAVIAQWNGNETGNSIADVGVDPNPSKLSLSPKRLQDN 662  
 Db 590 VVIYGSTPYITPWKVKNPAIVTAHGGQEGRAVSVDLGDVNPAGRLMMTWYLSANOL 649  
 QY 663 PAFINFRTEAGRTLYGHDVYVGYRYEFADKDVNFPFGHSLSTYTFASNLSSHDKGL 722  
 Db 650 PDMDY-----DIKGRKTYGFEDDVLVPRGHLSYARSYDHLQI-QKTNYL 697  
 QY 723 SVSLSTVNTGSGVPGAOVLAOLYVPLAKAKINRYPKELGFAKVELQPGETKAV--TIEQ 780  
 Db 698 TVSATYITNTSHLQEEVYQLYVR-CQTSRVKRLPLTKLGFKRLMKIRKGEQKVYTFLEPE 756  
 QY 781 EKYVAAYFDEERDQMCVEKGYEYVSDS---AAKGYALRG 821  
 Db 757 E---LAIWDVYTRERYCYEGQMYTIMGRSSKDIQLEETIELEGE 797

## RESULT 13

beta-xylosidase [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 27-Nov-2001  
 C:Accession: G90484  
 R:She, Q.; Singu, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch  
 JONG, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder  
 arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: G90484  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-754 <KUR>  
 A:Cross-references: GB:AE006641; MID:g13816430; PIDN:AAK43134.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SSO3032  
 C:Superfamily: Beta-glucosidase

Query Match 17.7%; Score 776.5; DB 2; Length 754;  
 Best Local Similarity 26.4%; Pred. No. 7.4e-45;  
 Matches 242; Conservative 129; Mismatches 271; Indels 273; Gaps 29;

QY 6 VEAIIKLTLAEKV-----DLAIGIDFWHTALP--KHGVSRLRTPDG-----PN 48  
 Db 4 ISLNLQMSIEEKIQAOLAIPIDALMEGRKFESEKARXYLKGIOITRVAGSRLGLPK 63  
 QY 49 GVRGTFFN-----GVPAACFPCTSLGTFNQTILEE 81  
 Db 64 EV--VKLVKVKQKLVENTRLKIPATIHIEECLSGLMGSSPAFPAIGLASTWPELTLN 121  
 QY 82 AGKMGKRAIAKSAHVILGPTINMQRSPFGGRFESIEDPFLAGLAALIRGIG--STG 140  
 Db 122 VASTIRSQGRILGVNQCISPLDYVCRDRPMGRCEETVEDDYIVASKGLAITGLQGETQ 181  
 QY 141 VQATIKHFLCND-QODRRAMQOSIYTERALBEIYALPQIAVRDSQCAFMTAYNGINGV 199  
 Db 182 LVATKAKHFAAHFPPGGNNIAQVHVGNRLETFEPEVAVVKIGKVMIMPAYHEIDGV 241  
 QY 200 SCSENPXYLDMRLKEMWMDGLIMSDYG-----YTSITEAVV---AAGLDEMP 245  
 Db 242 PCHGNPQTLITNLROEMGFQDIVVSDYDGTQRLQALHAKVVASNKKMAAILALESGVDIEFP 301



Db 12 EVRVDDLRSMTLEEKVVAQLGSV--WGVELIDRGRKFSREKAKELLKNGIGQITRPGSGT 69

QY 49 GV---RGTFNFN-----GVPA---AC-----PPGSGISLGSFTNQT 78

Db 70 NLEPQEAALVELVIOEFLVEETRGLIPAMIEHECLTGMGLSGTGNFPQALAMASTWDPDL 129

QY 79 LEEAGMMKKEKAIKASAHYILGPTINMORSPLGREGFESIGDEPPFLAGLGAALIRGIG- 137

Db 130 IERKMTTAVEDARKIGAHGCLAPVLVDAPDRNGKRIEETFGESPILYARBGVSIVKGLQG 189

QY 138 ---STGVQATIKHFL-CNDQEDRRMVQSTVTERALREIYALPFOJLARDSQCAFMTAY 193

Db 190 EDIKKGVAIVTKHFAGYSASEGGKNNAPTNIPEREKEVFLPFEEAAVKRANVLSVMSXY 249

QY 194 NGINVSCEGNPKYILDMARKKEMGDLMSDYG-----1YSTTEAVNAG 239

Db 250 SEIDGPPCANRKLTLDIRKMGFEGIVSDFAVKVEDYHRIARKDSEAKRLALEAG 309

QY 240 LLEMPGPRFGEETLKFWNSCKPFIHVIDQKAREVLOVKKCAASGVVENPSETVNN 299

Db 310 IYVELPKTCTY--QYIKDLVEKGIIEALIDEAVTVLRK---FMILGLEN-PIYVEVK 363

QY 300 T--PETALLRRVNGEIVLKNENNVLPLSKKKKTLLVGNPKAKQATYHGGSGSALRAY 357

Db 364 AKTESHROIALERAKSIILKND-GILPLQKKKVALIIPNAGE----- 407

QY 368 AVTPEDGLSKOLETPPSTYVGATTPPILGEOCLLPDCAFGKRWKRVFNPPGTPNQH 417

Db 408 -----VRNLLGDVWYL-----AHI 421

QY 418 DELFTKTDMHLYDYHHPKADTWYADMEGTAYADEDCYELGLVVCGRAKAYDDQLV 477

Db 422 RALLDNIDV-----FQNPQIPRNTYERLKS-----I 449

QY 478 DNATQVPS--DAFGSARRETGRINLVKGNTRYKKEIEFGSAPTYTLKGDITVPGHSL 535

Db 450 EEHMKSIPSLDAF-----KEEG-----IEFEYAK----- 474

QY 536 RKGSGKVI--DDQAEIKESVALAKKEHDVITC---AGLANDMFT-EGADASAKMLPGVLD 589

Db 475 ---GGEVIGEDSGEEALEIAKKSVAIVYVGDKSGLLTDCITGESRDAMNKLPLPVOE 531

QY 590 QLIADVAANPNTVVVMQGTPEEMP-WLDATPAVIVQAWYGNGETGNSIADVYFGDNP 648

Db 532 ELVLEAKKGRVYLVLLITGRPSLKNVYDKVALLIQWMLPGABGAGAIYDIIYKGNPS 591

QY 649 GKLSISFPRLODNPALFNRTEBAGRTLYGEDVYVGRYIEFADKDVNFPFGSLSTTF 708

Db 592 GKLPISFPRSAQOIPVFEHVKPSGGSHMGD-----YVDESTKPL-PPFGGLSTYKF 644

QY 709 AFSNLSVSHKD---GKLSVLSVKNTGVSVPQOVQOLYKPLQAKKINNPVELKGEAF 764

Db 645 EYSNRIERKEKPPPACEVVIKVDVENIGDRDGEVQOLI-GEFASVTPZYVELKGEFR 703

QY 765 VELQGETKAVTIEBQEKVAAVFDEERD-QWCVEKGEDEVIVSDSSAAKDGVALRGKFT 823

Db 704 VSLKAKEKTVFRLMDVLAY---NRDMKLVEEGEFKVMVGSSS---EDIRLTGSFS 757

QY 824 -VGE 826

Db 758 VVGE 761

RESULT 15

B45956

beta-glucosidase (EC 3.2.1.21) 2 precursor - yeast (Saccharomycopsis fibuligera)

C:Species: Saccharomycopsis fibuligera

CjDate: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 24-Sep-1999

CjAccession: B45956

R:Machine, M.: Ohtsuki, I.; Fukui, S.; Yamashita, I.

Appl. Environ. Microbiol. 54, 3147-3155, 1988

File: Nucleotide sequences of Saccharomycopsis fibuligera genes for extracellular

A:Reference number: A45956; MUID:89133518; PMID:3146949  
 A:Accession: B45956  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-880 <MAC>  
 A:Cross-references: GB:M22476; NID:9170809; PID:AAA34315.1; PID:9170810  
 C:Superfamily: beta-glucosidase 1  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 16.4%; Score 722; DB 2; Length 880;  
 Best Local Similarity 24.5%; Pred. No. 5e-41;  
 Matches 243; Conservative 118; Mismatches 267; Indels 362; Gaps 33;

```

QY 7 EALIKLTLAEKVDLAGIDFW-----HTKALPKHGVSLRFTDGPNGVGRGKFEKNGV 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 KALVSMITVEKKNLTGTG-WQLGPCVNTGSPVRFGIPNLCLODGLGRLTDEFTG- 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 PAACPCGTSLSSTENOTLLEAGKMKKEAIKSAHVILGPTINMRSPL-----GGRG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 ---YPSGMATGTFNKDLFQRGQALGHEFNKGVHIALGPV---GPLGVKARGRN 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 FESIGDPPFLAGAALRGIOSTGVQATIKHFLCNDQEDRRM----- 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 FEAFGSDPYLOGIAAATTKGLOENNVMAVCVHFIENEDIIYKOPSNSKVDEPDATKE 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 MVQSIATERALREIYALPQIAVRDSQPGAFMTAYNGINGVSCSENPXYLDGMLRKEMGW 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 SISANIPDRMHLYLMPFADSR-AGVSYVMSYRNVNNTYSCENSYMINHLKEELGF 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 DGLIMSDWTGYSTTEAVVAGLDLEMPG-----PRRGETLKENVSNCKPFIHVID 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 QGFVVDMAAQMGSAGSAISGLDMSMPGELLGWMNTGKSYGONLTKAVYNETVPIERLD 352
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 QRAREVL-----QFVKCAASG--VTENGPETTVNN-----TP 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 DMAFTRILALYATNSPPTKDLRPNSSFTTKEYGNEFFVDKTSPLYVKVNHVYDPSNDPTE 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 ETAAALRYKNGEIVLKNENNVLPSS--KKKTKLI---VGPNAK-----QAT 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 DFLAL---KVAEESIVLKKNEKNLPISPNKKVRLILSLGIAAGDPKGYCSDQSCVDGAL 469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 YHGGSSAALRAY--YAVTPEDGLSKOLETPPSTYTGAVTTPPIIGEOCLTPDGAPGRW 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 FEGWGGSGSV-GYPKYQVTFEEISA----- 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 403 RVENEPGPGPNRHOIDELEFTKDMHLVDYHHPKADTWYADMEGTYTADDECYELGLV 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 ---NAKKNKMQFDYIRESF-----DLTOYSTVASDHHMSIVV 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 463 VCGTARAYVDOLVVDNATKQVPGDAFGSATREETGRINLVKNTYKFKIEFGSAPTYT 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 SAVSGEGY---LIID-----GN----- 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 523 LKGDITVPGHGLRVGGCKVYIDDAIEIKSVLAKEHDOYIICAGLADMETEGADRA$M 582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 -RGD-----KNNVTL-----WHNS----- 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 KLRGVLDOLADVAANPNTVYVM-OTGTPPEEPMLDATPAVLAQWYG--GNETGNSIAD 639
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 -----DNLKAVAEKANTVYVITSTGOVDVESFADHPRVTAIYVAGPLGDRSGTALAN 609
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 640 VVEGDYNP$GKLSL$PPKRLQDNPAPLNFRTENG---RTLYGEDVYVGYRYEFADKDV 695
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 ILFGNANP$GHL$PFTVA$KNDYIPIVYTNPRGEPEDNTLAEHDLVDYRYFEKNI$P 669
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 696 NPF$GGLSTYTF$F$NLSV$H----- 717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 RYAF$GLSYNEYK$VNAK$VDEELPQKLYLAEYSYNKTEEIINNPEDAFF$P$NA 729
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 718 -----KDGK-----LSVLSV 728
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 730 RRIQ$FLPYLD$NVTLKDGNT$EYPDGYST$EORTTPIQPGGIGLGDALMEVAYK$EVDY 789
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 729 KNTG$VPG$QV$AOLYV$P$Q$AKINRPYKELKGF$AKV$ELOPGETKAVTIEO$EKY$AA$YF 788
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 790 QNLG$NSTDK$EVP$OLY$KHP$EDGK$FET$PV-QLRG$PEKVEL$P$G$EKKY$E$F$EL$RDL$SVM- 847
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 789 DEERDQK$VEK$G$DY$E$V$Y$D$S$A$K$D$V$AL 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 848 DTTROS$W$VES$GTY$E$AL-----GVAV 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: April 26, 2003, 13:10:42  
 Job time : 56 secs